Preface

Information systems are able to deal with an incredible amount of data, concerning every aspect of real life. Data are dealt with by heterogeneous systems, concerning heterogeneous application domains, are structured in heterogeneous data formats; furthermore, semi-structured (XML) or unstructured (textual) documents are becoming important data sources for enterprises and organizations.

Enterprises and organizations need to deal with such heterogeneous and often very large volumes of data, which may also be uncertain, imprecise, and incomplete. For instance, a typical question is “How do we take advantage of such large (and possibly heterogeneous) volumes of data, in order to extract useful information for decision making, for discovering and sharing useful knowledge, etc.?” In other words, they need technology for *Flexible Data Management in Databases and Information Systems*.

Flexibility in such a context might be intended in several ways; some of them might be the following (but others may be found):

- the capability of easily querying large volumes of data, in order to obtain significant or useful descriptions;
- the capability of integrating heterogeneous data, possibly inconsistent, coming from different, distributed, and incoherent data sources;
- the capability of integrating structured, semi-structured, and unstructured (textual) data;
- the capability of modelling databases, defining query languages, building information systems dealing with large volumes of data, heterogeneous and distributed data, uncertain or imprecise data, changing in time data.

Researchers working on data models and techniques able to manage imprecise and uncertain data in databases are proposing database models that are able to deal with imprecision and uncertainty in the data modelling phase, as well as during querying. They generally come either from the database community or from the soft computing community, and are separately dealing with the topic. Researchers working on data mining and knowledge discovery techniques, are providing technology and tools for discovering useful and/or unexpected information from large volumes of data.

The information retrieval community, which is providing solutions for Web search engines, is now moving toward new emerging research themes, such as the definition of more effective techniques to index, categorize, filter, and retrieve collections of multimedia documents heterogeneously structured and rapidly changing in time, problems that are faced by the database and data mining communities as well.

This initiative is the natural evolution of successful SIUFDB-04 (DEXA 2004 International Workshop on Supporting Imprecision and Uncertainty in Flexible Databases) and IDDI-05 (DEXA 2005 International Workshop on Integrating Data Mining, Databases and Information Retrieval), and like them it is aimed at yielding a cross-fertilisation in the specific field.

The workshop opens with the invited paper written by S. Zadrozny and J. Kacprzyk, titled “Bipolar Queries and Queries with Preferences.” The authors show how it is possible to formalize in a unified way the concept of “bipolar queries” and the concept of “query with preferences.” The paper shows that both the approaches followed by Chomicki (from the database community), that defined a solution for the concept of “queries with preferences,” and the approach followed by Zadrozny, Kacprzyk, and other researchers (from the soft computing community), that defined the concept of “bipolar queries,” are in effect two faces of the same problem and are based on the same formal basis.

The Flexibility in Databases session includes three papers concerning the problem of adding flexibility to databases. The paper titled “A Knowledge-Based Approach for Database Flexible Querying,” by N. Hachani and H. Ounelli, proposes a technique to provide approximated answers to a database query when there is no exact answer. The paper “Avenues to Flexible Data Integrity Checking,” by H. Decker and D. Martinenghi, introduces the concept of tolerant integrity constraint: it extends the classical notion of integrity constraint, so that the degree of inconsistency in a database not
satisfying the integrity constraints does not increase from one database state to the next one. The paper “On a Qualitative Approximate Inclusion—Application to the Division of Fuzzy Relations,” by P. Bosc and O. Pivert, proposes a relaxed definition of “inclusion” that copes with the presence of unsatisfied integrity constraints in the field of relational databases.

The Data Mining sessions includes two papers. The first one, “Using Gaussians Functions to Improve Clustering Prototypes” by M. Sassi, A. Grissa Touzi, and H. Ounelli, presents a clustering technique based on the evaluation of fuzzy measures that allow the algorithm (presented in the paper) to define the cluster centers. The paper “dmFSQL: A Language for Data Mining,” by R. A. Carrasco, M. A. Vila, and F. Araque, presents an SQL-like language to specify data mining tasks based on clustering and classification.

In the Demonstration session, G. Pasi, G. Bordogna, and R. Villa present a demonstration of the PENGI system, a system for information retrieval in news that they designed and developed in the context of a EU-funded project. The proceedings include a paper describing the system, while the authors demonstrate the system at the workshop.

Two papers are included in the final session, Searching and Context-Aware Computing. The paper “MetaOn—Ontology Driven Metadata Construction and Management for Intelligent Search in Text and Image Collections,” by H. Karanikas, N. Peleakis, D. K. Ikovidis, I. Kopanakis, T. Mavroudakis, Y. Theodoridis, presents the basic idea of the MetaOn project, funded by the European Union and the Greek government, whose goal is to develop theory, techniques, and a prototype system for metadata management, knowledge discovery and delivery, based on image collections. The final paper, “An Evaluation of Query Languages for Context-Aware Computing” by P. D. Haghighi, A. Zaslavsky, and S. Krishnaswamy, presents an evaluation of query languages for context-aware computing; a list of features characterizing these languages is presented and, based on it, several languages are compared and evaluated.